IN THE SPECIFICATION:

Please REPLACE the paragraph beginning at page 15, line 25 with the following paragraph:

Preferably, the hash value for the filtered document is determined using a hash algorithm having an approximately even distribution of hash values. More preferably, the hash value for the filtered document is determined using a secure hash algorithm. With a secure hash algorithm, the probability of two token streams creating the same hash value is reduced. Even more preferably, the hash value for the filtered document is determined using the hash algorithm SHA-1. With the hash algorithm SHA-1, each retained token in the arranged token stream is processed individually to obtain a hash value for the filtered document. With the SHA-1 hash algorithm, which uses 160 bits, the probability of duplicate values for different token streams is $\cancel{O}\underline{P}(2^{-160})$.